metin, alet, tasarım, cebir içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Report**

An Analysis on Factors Affecting Drug Abuse and its Consequenses

by:

Gökhan VAROL

Adil Emre KALKAN  
Çağatay OKTAY

Ömer KOLCU

2025 January

METU/ODTÜ  
Orçun Arda RENDA

Ata KAVAK

Abstract

This study investigates the relationship between drug usage and various factors such as gender, economic status, age, and mental health. The dataset, initially messy and incomplete, underwent significant cleaning and tidying to ensure the accuracy of the analysis. Using Python and the seaborn library, we visualized key trends and patterns, focusing on how drug type influences mental health across genders, the impact of economic status on monthly drug expenditures, and the relationship between drug usage frequency and economic status. The findings provide valuable insights into the complex factors affecting substance use and its consequences.

# Introduction

Drug addictions are a serious concern regarding young adults’ health in today’s world. Understanding what factors affect these addictions and to what extent they take toll on the lives of young adults is crucial to diagnose and to prevent drug abuse. This report used a dataset providing information about the social, economic and health conditions of young adults aged 15 – 27 as well as details about their drug use. The dataset was first cleaned and tidied with NumPy and Pandas libraries in Python, using appropriate methods to be prepared for the analysis. After the inspection of the cleaned dataset, research questions were formulated and multivariate correlations were sought in plots and charts with Matplotlib and Seaborn libraries in Python. The aim of this analysis is to reveal factors causing/supporting drug abuse and the effect of drug addictions on the lives of young adults.

# Data Description

Before cleaning the dataset, it contained 735 entries and 20 variables. 8 of these variables are numerical and 12 of them are categorical. The variables are described as followed:

1. **Year:** The year of the observation (2000-2023).

*(Numerical – Discrete)*

1. **Age:** Age of the individual (15-27).

*(Numerical – Discrete)*

1. **Gender:** Gender of the individual (Male, Female, Non-binary).

*(Categorical – Nominal)*

1. **Drug Type:** Type of drug used (Marijuana, Alcohol, Cocaine, Prescription, Hallucinogens, Others).

*(Categorical – Nominal)*

1. **Frequency:** Frequency of drug usage (Daily, Weekly, Monthly, Rarely).

*(Categorical – Ordinal)*

1. **Peer Influence**: Level of peer influence (High, Medium, Low).

*(Categorical – Ordinal)*

1. **Family History:** Whether the individual has a family history of substance abuse (Yes, No).

*(Categorical – Nominal)*

1. **Mental Health Issue:** Presence of mental health issues (Yes, No).

*(Categorical – Nominal)*

1. **School Performance:** Academic performance (Excellent, Good, Average, Poor).

*(Categorical – Ordinal)*

1. **Economic Status:** Economic background (Low, Middle, High).

*(Categorical – Ordinal)*

1. **Region:** Geographic region of residence (Urban, Suburban, Rural).

*(Categorical – Nominal)*

1. **Access To Treatment:** Ease of access to treatment (Easy, Moderate, Difficult).

*(Categorical – Ordinal)*

1. **Parental Supervision:** Level of parental supervision (High, Moderate, Low).

*(Categorical – Ordinal)*

1. **Criminal Record:** Whether the individual has a criminal record (Yes, No).

*(Categorical – Nominal)*

1. **Usage Duration Years:** Duration of drug use in years (1-9).

*(Numerical – Discrete)*

1. **Monthly Expense Drug:** Monthly expense on drugs in USD (50-500).

*(Numerical – Continuous)*

1. **Mental Health Score:** Mental health score on a scale of 1 to 100.

*(Numerical – Discrete)*

1. **Hours Spent Socializing Per Week:** Weekly hours spent socializing (0-40).

*(Numerical – Continuous)*

1. **Annual Income:** Annual income in USD (1,000-50,000).

*(Numerical – Continuous)*

1. **Number of Friends Using Drugs:** Number of friends involved in drug use (0-20).

*(Numerical – Discrete)*

# Data Cleaning and Tidying Steps

To clean and tidy the data, dataset was first uploaded to Google Colab as it makes reaching and using Python libraries easier and the data was inspected with .info(), .head() and .tail() functions. Within the inspection, it was found that the column names had typos; entries had unnecessary spaces, unordered use of uppercase and lowercase letters, arbitrary symbols, missing values and negative integer values in the “Number of Friends Using Drugs” column where every entry should be non-negative. Hence after spotting the issues, the column names were first made readable using camel case convention. Variables with ‘object’ data type entries were stripped from unnecessary spaces and signs, and all characters were lowercased. The negative values were handled by converting them to absolute values. ‘Annual Income’ column was marked as ‘object’ data type as the entries contained ‘$’ sign. After removing the sign, variable’s data type was converted to int64. Later, missing values were handled by replacing these values with the mean in numerical columns and by replacing them with the mode in categorical columns. Next, the duplicates were dropped and descriptive statistics of the data were checked to see if there are any unusual outliers. An extreme outlier was found in ‘Number of Friends Using Drugs’ column and it was replaced by the mean. Lastly, the variables 'Year', 'Age', 'Usage Duration Years', 'Mental Health Score', 'Annual Income', 'Number of Friends Using Drugs' were all converted to int64 data type, as the missing values replaced by the mean automatically converted them to float64 data type. With this last step, dataset was found to contain 732 entries with two ‘float64’, six ‘int64’ and twelve ‘object’ columns and was ready to be used in visualizations.

# Descriptive Statistics

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Year | Age | Usage Duration Years | Monthly Expense Drug | Mental Health Score | Hours Socializing Per Week | Annual Income | Number of Friends Using Drugs |
| count | 732.0 | 732.0 | 732.0 | 732.0 | 732.0 | 732.0 | 732.0 | 732.0 |
| mean | 2011.887978 | 20.821038 | 4.811475 | 282.358951 | 48.318306 | 19.601067 | 24084.598361 | 9.273224 |
| std | 6.547189 | 3.607553 | 2.545881 | 126.267644 | 27.756683 | 11.310075 | 14387.565501 | 5.586939 |
| min | 2000.000000 | 15.000000 | 1.000000 | 50.110000 | 1.000000 | 0.200000 | 1161.000000 | 0.000000 |
| Q1 | 2007.000000 | 18.000000 | 3.000000 | 184.480000 | 25.000000 | 10.265000 | 10435.000000 | 4.000000 |
| Q2 | 2011.000000 | 20.000000 | 5.000000 | 282.158928 | 48.000000 | 19.595457 | 23987.000000 | 10.000000 |
| Q3 | 2018.000000 | 24.000000 | 7.000000 | 385.645000 | 71.250000 | 28.660000 | 36774.000000 | 14.000000 |
| max | 2023.000000 | 27.000000 | 9.000000 | 499.710000 | 99.000000 | 39.990000 | 49984.000000 | 19.000000 |

# Explanatory Data Analysis

Before hypothesizing research questions and creating visualizations useful for the analysis, multivariate correlations between different variables were checked with different plots and charts such as ordered bar charts, box plots and violin plots to find relationships between categorical and numerical variables. After this initial analysis, appropriate questions were asked and visualizations were made accordingly.

**Research Question 1: How drug type affects mental health by gender? (Gökhan Varol)**

The purpose of this question is to explore the relationship between drug use and mental health, focusing on how this relationship differs across genders: female, male, and nonbinary. Mental health is a complex and multifaceted issue, and drug use is known to be a significant factor influencing mental well-being. By examining various drug types and their impacts, this question aims to identify patterns and provide a deeper understanding of these interactions, particularly when considering gender-specific differences. **metin, diyagram, ekran görüntüsü, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu**

*Visualization1: Drug Type vs Mental Health Score by Gender*

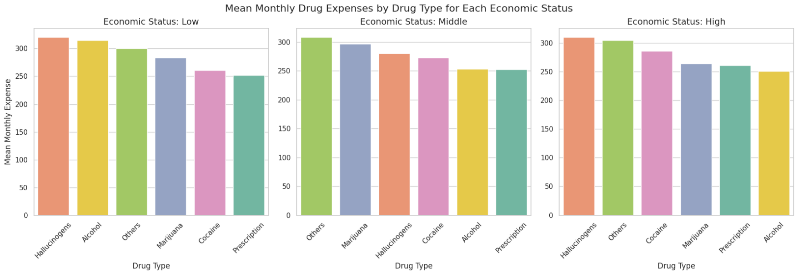
The analysis focuses on six main drug categories: prescription drugs, hallucinogens, marijuana, cocaine, alcohol, and other substances. Each category demonstrates unique effects on mental health. For instance, prescription drugs appear to have a more positive influence on mental health scores among male users, while females and nonbinary individuals tend to have slightly lower scores in this category. Hallucinogens show a similar pattern of impact across all genders, with a wide distribution of mental health scores, suggesting no significant gender-based variation. Marijuana shows notable variability, with nonbinary users tending to have slightly lower scores compared to male and female users. Cocaine, on the other hand, appears to have a more negative impact on mental health for nonbinary individuals, while males and females exhibit similar patterns. Alcohol has a relatively neutral effect across all genders, with less variability in scores compared to other drug types. Lastly, the "others" category does not show pronounced gender differences, with middle values remaining consistent across groups.

When examining gender differences, certain trends become apparent. Male users generally report higher mental health scores for prescription drugs and alcohol compared to females and nonbinary individuals. Female users, however, tend to exhibit lower mental health scores in response to cocaine and prescription drugs. Nonbinary users often report the lowest mental health scores across most drug types, which could be attributed to a combination of drug effects and external social or psychological factors. These patterns highlight the importance of considering gender as a critical variable when studying the relationship between drug use and mental health.

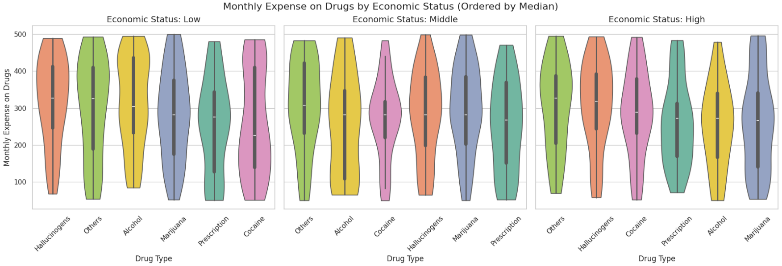
In conclusion, this study underscores the complexity of the interaction between drug use and mental health, demonstrating how gender plays a pivotal role in shaping these outcomes.

**Research Question 2: How does economic status and each type of drug affect monthly expense on drugs? (Çağatay Oktay)**

This question aims to find if there is a correlation between the economic status of the individuals and their expenses on different types of drugs. A clear answer to this question can provide valuable insights on which drugs are more common in different economic classes and can be used to combat drug abuse more effectively. To provide an answer to this question, first a lattice of ordered bar charts was plotted for each economic status demonstrating mean monthly expense on drug types; then to be able to see the distributions of each drug type more clearly, comparative violin plots were plotted:



*Visualization2: Mean Monthly Expenses by Drug Type for Each Economic Status*



*Visualization2.1: Monthly Expense on Drugs by Economic Status (Ordered by Medians)*

From the bar charts, some relationships can be seen for drug types and economic status. It is seen that the mean monthly expense on alcohol decreases as economic status increases, and a converse relationship is seen with cocaine: mean monthly expense on cocaine increases as economic status increases. Low and high economic classes both spend the most on hallucinogens, while middle class spends the most on marijuana and other types of drugs. From the three bar charts, it is also seen that prescription drugs generally have the least expenses with only exception being alcohol for high economic class.

Additional inferences can be made from violin plots. It is seen that top 3 drug types (hallucinogens, alcohol and others) in low economic class all have negatively skewed distributions, indicating more frequent spendings above the median in this economic class. For middle and high classes, modes of distributions are closer to median: distributions in these classes are generally less variable compared to low economic class.

Therefore, from the two visualizations, it seen that for alcohol there is a negative correlation between the economic status and monthly expenses and for cocaine there is a positive correlation between the economic status and monthly expenses. Individuals from low economic status tend to have more extreme spendings on drugs while individuals from middle and high classes have more moderate spendings compared to each other.

**Research Question 3: What is the relationship between drug usage frequency, economic status, and regional differences? (Orçun Arda Renda)**

This question aims to explain the effects of economic status and places of residence on people's drug usage frequency. In the provided dataset, economic status is categorized as low, middle, and high, while places of residence are classified as rural, suburban, and urban. Within these classifications, drug usage frequency is compared as daily, weekly, monthly, and rarely, supported by various visualizations.

The results show that individuals with low economic status have significantly higher daily and weekly drug usage rates. This suggests that economic difficulties push people toward drug use. On the other hand, individuals with high economic status exhibit higher rarely usage rates. It has been observed that a higher level of welfare has a reducing effect on drug use. These findings indicate an inverse relationship between economic status and drug usage frequency; as welfare increases, usage frequency decreases.

In rural areas, despite the challenges of accessing drugs compared to urban centers, individuals with low economic status still show higher daily and weekly usage rates. This can be explained by the lack of access to various services and limited communication networks in rural areas. Suburban regions generally exhibit a balanced pattern, likely due to their intermediate position between urban and rural settings. In urban areas, monthly and rarely usage rates are higher compared to other regions. The easy access to drugs in cities may have contributed to the relatively higher usage frequency.

In conclusion, low economic status is strongly associated with more frequent drug use. Geographical conditions also influence the tendency toward drug use. Individuals in rural areas face greater risks compared to those in urban settings.

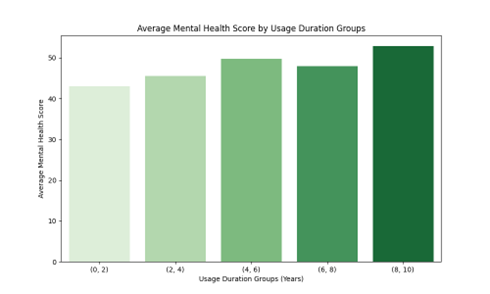
A graph of different colored bars

Description automatically generated with medium confidence

*Visualization3: Impact of Economic Status and Regional Differences on Drug Usage Frequency*

**Research Question 4:** **How the duration of substance use affects individuals' mental health and social behavior? (Adil Emre Kalkan)**

This question aims to explore how the duration of substance use affects individuals' mental health and social behavior. Specifically, the objective is to analyze whether longer durations of usage correlate with changes in mental well-being and weekly social engagement. To address these questions, two bar charts were created based on the data. The first chart illustrates the average mental health scores across different usage duration groups.



*Visualization4.1: Average Mental Health Score by Usage Duration Groups*

The second chart examines the average hours spent socializing per week within the same groups. Contrary to the first chart, it reveals a gradual decline in socializing hours as usage duration increases, with a significant decrease observed in the 8–10 years group. This decline indicates that prolonged substance use may have a detrimental impact on social interaction and engagement.metin, ekran görüntüsü, dikdörtgen, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

*Visualization4.2: Average Hours Socializing Per Week by Usage Duration Groups*

Overall, the analysis reveals that the duration of substance use has mixed effects on individuals. Longer usage appears to correlate with higher mental health scores, potentially indicating improved coping mechanisms. However, it also shows a noticeable decline in socializing hours, suggesting a negative impact on social interactions. These findings provide a nuanced perspective on the relationship between substance use, mental health, and social behavior.

**Research Question 5: What is the relationship between age and mental health score by alcohol, cocaine and marijuana? (Ömer Kolcu)**

This question investigates the relationship between age and mental health scores, focusing on the use of three specific substances: marijuana, cocaine, and alcohol. The primary objective of this research is to understand how substance use impacts mental health across different age groups, providing insights into potential intervention points for improving public health outcomes.

metin, diyagram, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu

*Visualization5: Age vs Mental Health Score by Drug Type*

The line chart visualizes how mental health scores vary with age for individuals using marijuana, cocaine, and alcohol. The results show distinct trends for each substance. For marijuana users, mental health scores appear relatively stable across most age groups, with slight fluctuations in younger and older ages. Cocaine users, on the other hand, exhibit a significant decline in mental health scores as age increases, suggesting that prolonged use may exacerbate mental health challenges. Alcohol users demonstrate a more moderate trend, with mental health scores slightly decreasing over time, though not as drastically as observed with cocaine. The findings reveal distinct patterns in the relationship between age and mental health scores for users of marijuana, cocaine, and alcohol. While marijuana users exhibit relatively stable mental health scores across age groups, cocaine use is associated with a sharp decline in mental health with increasing age. Alcohol users show a moderate decline over time. These results highlight the varying impacts of different substances on mental health and underscore the need for further research to understand the underlying mechanisms driving these trends.

**Research Question 6: What is the relation ship between socializing hours and mental health score ?(Ata Kavak )**

This question aims to understand relationship between individuals socializing hours and mental health score. The main goal of this research is to

observe the how peoples socialing affect their mental healthscore.

**A graph with a red line

Description automatically generated***Visualization6: Impact of Socializing to the Peoples Mental Health Score*

According to scatter plot, we can say that there is a weak correlation as the regression line is similar to a slightly inclined line which means there is tiny proportional change among them. As the slight slope and the spread of data points offers a weak correlation .Thus ,socializing hours may not strongly predict a mental health score.

The shaded red area around regression line shows the confidence interval. The width of the area indicates the uncertainty situation in the trend.

# Conclusion

In conclusion, our analysis highlighted several significant relationships between drug usage and various socio-economic factors. Drug types were shown to have varying effects on mental health depending on gender, with notable differences in how individuals of different economic backgrounds spend on substances. The frequency of drug use was found to be closely linked with economic status, emphasizing the role of financial factors in substance use patterns. Additionally, the duration of drug use was shown to impact not only mental health but also social behavior, pointing to the long-term consequences of substance abuse. Finally, age-specific mental health scores revealed significant variations across substances, with alcohol, cocaine, and marijuana each having distinct effects. These findings contribute to a better understanding of the multi-dimensional factors influencing drug use and its outcomes.